Air Pollution Consequences in São Paulo: Evidence for Health

Bruna Guidetti

IPE/USP

FAPESP

Summary

- **Objective**: Investigating the impacts of air pollution on hospitalizations due to respiratory disease in São Paulo Metropolitan Area.
- **Motivation:**
  - Pollutants negatively impact human health, especially of vulnerable groups such as children and elderly.
  - There are few evidences for developing countries.
  - Frequent episodes of poor air quality in SPMA

Summary

- **Problem:** Endogeneity of air pollution exposure.
- **Solution:** Instrumental variables (wind variables)
- **Data:**
  - Air Pollution: São Paulo Environmental Company (CETESB)
  - DATASUS: daily hospitalizations due to respiratory disease.
- **Economic Literature**: Currie and Neidell (2005), Chay and Greenstone (2003), Neidell (2004), Lewis and Severnini (2015), Hanna and Oliva (2015), Chagas et al. (2016), Schlenker and Walker (2016).

Endogeneity Problem

- **Pollutants are not randomly allocated**
- **Avoidance behavior:** Neidell(2004) discusses that individuals might avoid activities that expose them to air pollution, in order to reduce negative externalities.
- **Economic activity :** the level of economic activity, which is positively correlated with air pollution, may cause a negative bias on the pollution impacts on health by income increase (Hanna and Oliva (2015); Herrnstadt e Muehlegger (2015))

Endogeneity Problem

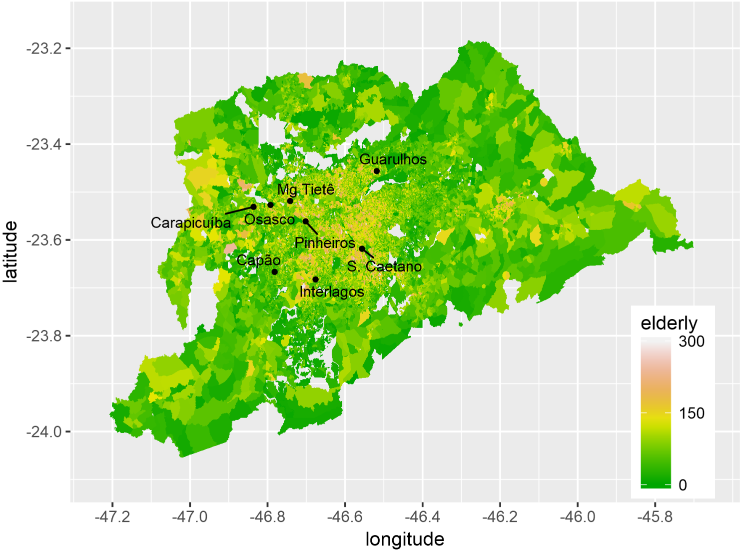
Strategies to deal with endogeneity

- Neidell (2004): amount of smog alerts.
- Chay and Greenstone (2003): *Clean Air Act Amendments* (CAAA).
- Chay and Greenstone (2003): economic recession in United States between 1980 and 1982.
- Hanna and Oliva (2015): closure of an oil refinery in Mexico City Metropolitan Region.
- Herrnstadt and Muehlegger (2015): wind speed and direction.
- Schlenker and Walker (2016): airport congestion in California

Data construction
CETESB data

- Instrument: wind speed
- Pollutant: NOx (ppb)
- Unit of observation: 8 monitors throughout SPMA from January to June in 2013, on a daily basis.
- Hospitalizations: number of **elderly** (aged 60 or above) hospitalized due to respiratory disease living within 5km radius around each of the 8 monitors.
- Dependent variable: hospitalization rate per 100,000 elderly.

Monitors



-23.2
-23.4
-23.6
-23.8
-24.0
latitude
-47.2
-47.0
-46.8
-46.6
-46.4
-46.2
-46.0
-45.8
longitude
Guarulhos
Mg Tietê
Carapicuíba
Osasco
Pinheiros
S. Caetano
Capão
Interlagos
elderly
300
150
0

Descriptive Statistics

Table: General Characteristics of the Monitors

	daily hospitalizations						
Monitors	**mean**	**minimum**	**maximum**	**total**	**average hospitalization rate**	**elderly population**	**average NOx**
Capão Redondo	3,48	0	10	629	4.00	86,919	25.55
Carapicuíba	1.04	0	5	188	2.81	36,932	39.10
Interlagos	3.88	0	12	702	4.28	90,655	29.08
Marginal Tietê	1,25	0	4	226	2.12	58,883	105.19
Osasco	1.28	0	6	232	2.40	53,484	84.67
Guarulhos - Paço Municipal	1.83	0	6	332	2.80	65,576	26.24
Pinheiros	0.99	0	4	179	0.87	114,003	69.47
São Caetano do Sul	4.27	0	12	772	3.09	137,811	44.20

Source: DATASUS, CETESB and Census-2010

Specification

- **2 stages estimates:**

$$log(pollution_{it}) = \alpha + \beta_1 ws_{it} + \beta_2 ws_{it-1} + \theta_i + \mu_t + \epsilon_{it} \text{ (1st stage)}$$

$$rate_{it} = \gamma + \lambda log(\hat{pollution}_{it}) + \eta_i + \delta_t + \varepsilon_{it} \text{ (2nd stage)}$$

  - Wind speed:
    - scalar-based: speed average (m/s)
    - vector-based: speed weighted by wind direction

Identification Hypothesis

$$\mathbb{E}(\mathbf{z}_{it}\varepsilon_{it}/\eta_i, \delta_t) = 0, \text{ where } \mathbf{z}_{it} = (ws_{it}, ws_{it-1})$$

Results

Table: First Stage

	Dependent variable: $\log(NOx_{it})$		
	(1)	(2)	(3)
Scalar-based			
ws_{it}	-0.461***	-0.518***	-0.460***
	(0.072)	(0.060)	(0.049)
ws_{it-1}	-0.052	-0.114***	-0.040
	(0.037)	(0.038)	(0.043)
F	23.164	57.350	50.930
Sargan (p-value)	0.199	0.363	0.387
Vector-based			
ws_{it}	-0.357***	-0.354***	-0.309***
	(0.047)	(0.044)	(0.037)
ws_{it-1}	-0.092**	-0.089***	-0.056***
	(0.033)	(0.018)	(0.017)
F	29.348	52.761	51.855
Sargan (p-value)	0.266	0.501	0.431
Monitor fixed effect	No	Yes	Yes
Time fixed effect	No	No	Yes
Observations	1267	1267	1267

Results

Table: Reduced Form

	Dependent variable: $rate_{it}$		
	(1)	(2)	(3)
Scalar-based			
ws_{it}	0.159	-0.257*	-0.232**
	(0.171)	(0.134)	(0.108)
ws_{it-1}	0.256	-0.154	-0.109
	(0.170)	(0.101)	(0.100)
Vector-based			
ws_{it}	0.053	-0.222***	-0.200***
	(0.110)	(0.055)	(0.045)
ws_{it-1}	0.148	-0.126	-0.116
	(0.126)	(0.103)	(0.096)
Monitor fixed effect	No	Yes	Yes
Time fixed effect	No	No	Yes
Observations	1267	1267	1267

Results

Table: Second Stage

	Dependent variable: $rate_{it}$		
	(1)	(2)	(3)
Scalar-based			
$log(NOx_{it})$	-0.710	0.606**	0.594**
	(0.462)	(0.242)	(0.235)
Vector-based			
$log(NOx_{it})$	-0.360	0.721***	0.752***
	(0.353)	(0.259)	(0.243)
Monitor fixed effect	No	Yes	Yes
Time fixed effect	No	No	Yes
Observations	1267	1267	1267

Results

Table: Regressions without instrumental variable

	Dependent variable: $rate_{it}$		
	(1)	(2)	(3)
$log(NOx_{it})$	-0.381**	0.387**	0.158
	(0.130)	(0.163)	(0.109)
Monitor fixed effect	No	Yes	Yes
Time fixed effect	No	No	Yes
Observations	1267	1267	1267

Robustness check

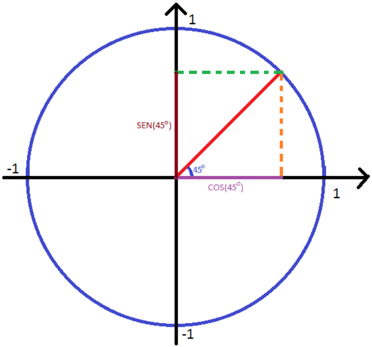
- First stage: regression of air pollution on wind speed registered days after the hospitalization.
- Reduced form: regression of hospitalization rate on the wind speed of another monitor (randomly chosen).
- Reduced form:regression of hospitalization rate for digestive system disease on wind speed.

Limitations

- Few monitors
- Missings
- No controls (such as temperature and humidity)
- Solution: INPE data

Velocidade do vento vetorial

Herrnstadt and Muehlegger (2015):

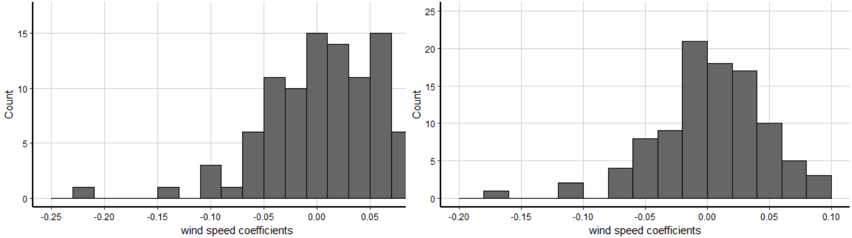


$$\vec{ws} = \sqrt{(sin(\theta) * ws)^2 + (cos(\theta) * ws)^2}$$

back

- Steps:
  - Finding $sin(\theta) * ws$ and $cos(\theta) * ws$
  - Calculating the hourly average for each day

Robustness check



(a) scalar-based (b) vector-based

Figure: First stage falsification

back

Robustness check

Table: Placebo I

	Dependent variable: $rate_{it}$	
	(1)	(2)
	Scalar-based	Vector-based
ws_{jt}	-0.147	-0.045
	(0.142)	(0.085)
ws_{jt-1}	-0.113	-0.082
	(0.075)	(0.053)
Monitor fixed effect	Yes	Yes
Time fixed effect	Yes	Yes
Observations	1267	1267

back

Robustness check

Table: Placebo II

	Dependent variable: $rate_{it}$	
	(1)	(2)
	Scalar-based	Vector-based
ws_{it}	-0.112	-0.060
	(0.093)	(0.073)
ws_{it-1}	0.220***	0.137**
	(0.063)	(0.058)
Monitor fixed effect	Yes	Yes
Time fixed effect	Yes	Yes
Observations	1267	1267

back